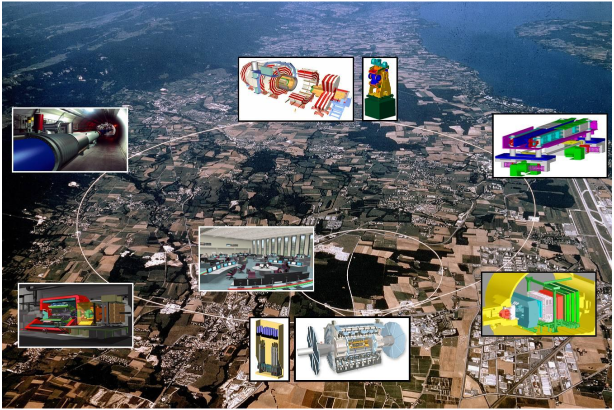
ML Applications for Particle Physics

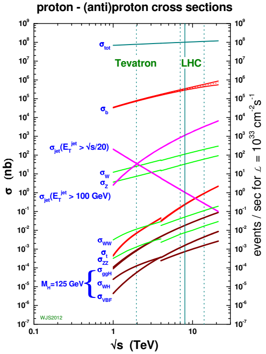
Ian J. Watson

University of Seoul

Artificial Intelligence & Machine Learning for Fundamental Science
November 15-17, 2021

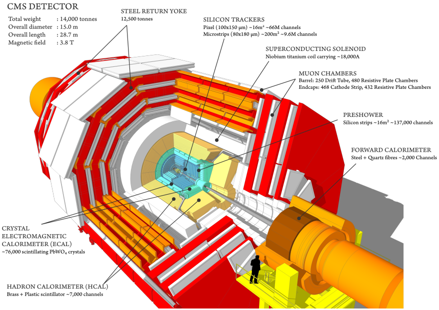
Context: LHC

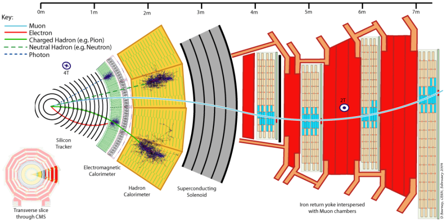




- 4 experiments on the LHC proton-proton collider ring: CMS, ATLAS, LHCb, ALICE
- At UoS we work at CMS: Bunch crossing rate of ~ 40 MHz, from which typically ~ 1 kHz stored to disk for further analysis
  - Each bunch crossing has dozens of "pileup" collisions
  - From this data haystack, searching for extremely rare processes

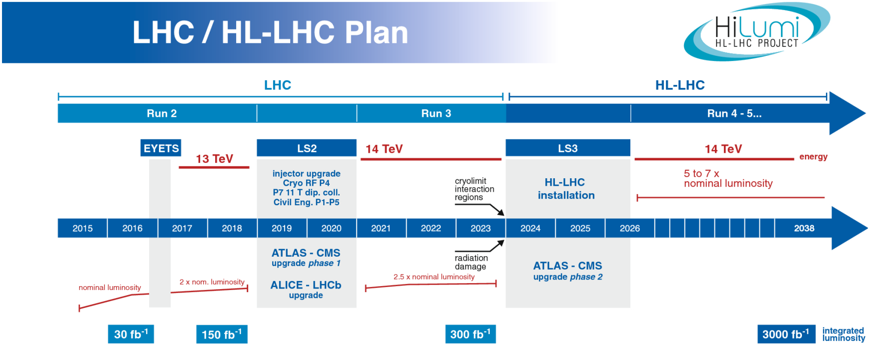
Context : CMS





- 14,000 tonnes, 28.7 meters long, 15 meters in diameter
- Instrumented with detectors arranged in an onion structure
  - Provides particle reconstruction and identification ($e^{\pm}/\gamma$ vs hadronic vs $\mu^{\pm}$), particle momentum and energy measurements
- Used to find the Higgs boson (along with ATLAS), probing the Standard Model of Particle Physics and searching for new physics beyond the Standard Model

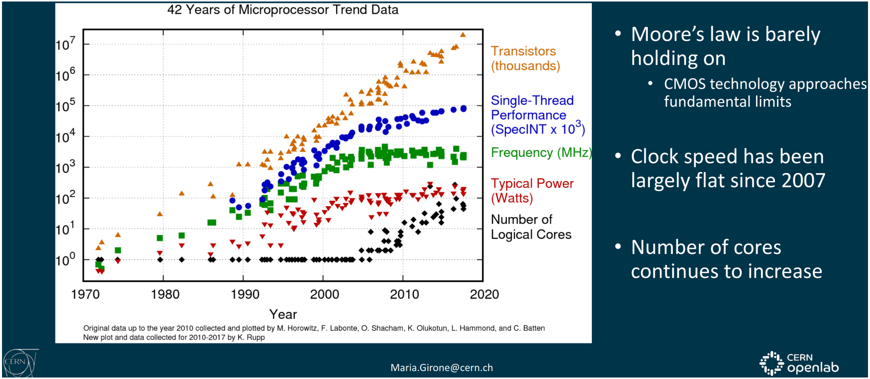
Context : HL-LHC



- Over the decades we willl increase luminosity
- Expect to collect about 30x more data than we currently have
  - Go from $\approx$ 500 PB of stored data total, to several EBs
- We will need a commensurate increase in the amount of simulation, ability of data acquistion, speed of user analysis, etc.
  - Could say similarly for e^+e^- (B2), neutrinos (DUNE), etc.
- HL-LHC potentially followed by FCC, takes the roadmap to the 2090s

Albrecht et al, "A Roadmap for HEP Software and Computing R&D for the 2020s", Computing and Software for Big Science (2019) 3:7

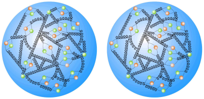
Outcome : Looking to industry tools, used with HEP Flavor

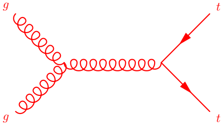


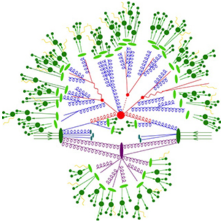
- FPGAs and GPUs starting to be used throughout the stack
  - The big gains from now come from using heterogenous computing
- Machine learning techniques increasingly used at all levels
- Take advantage of tools from outside HEP: Kubernetes for cluster management, docker, Tensorflow/Keras or PyTorch, Jupyterlab, etc.
- Talk today about some studies we are doing at UoS

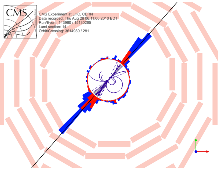
R&D: HEP Computing Infrastructure @ Update of the European Strategy for Particle Physics 2019

Events in particle physics





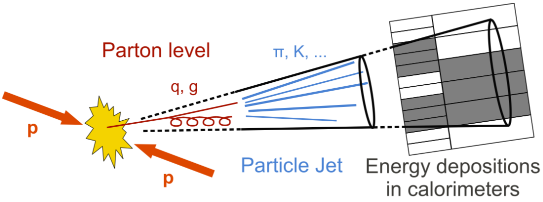


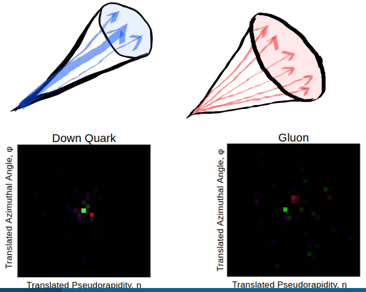


- Proton: two *u*, one *d* quark, bound together through the exchange of gluons, which can also split into temporary pairs of "sea quarks"
- Parton Level: Hard collision of *partons*
  - At high energies, the quarks and gluons inside the proton act as free particles
- Particle: Stable particles after parton evolution
  - Quarks and gluons evolve into **jets**, a collimated spay of particles
- Detector: The (semi-)stable final state particles propogate and interact with the detector
- Reconstruction: From the detector information, reconstruct the event
  - Combine hits in the silicon detectors to form charged particle tracks
  - *Cluster* energy deposits in calorimeter cells
  - Reduce into *reconstructed* e^-, γ, hadrons

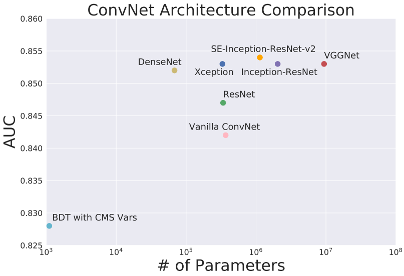
Quark vs Gluon Jets

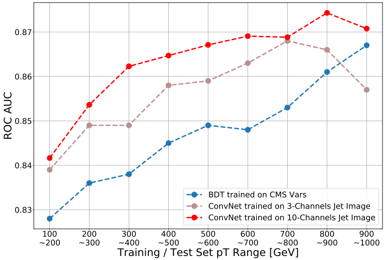
- When a high energy quark or gluon is produced in the pp collision, due to color confinement, we don't detect the it directly but rather the spray of hadrons that get produced by QCD processes, a *jet*
- Create an image by pixelizing the reconstructed tracks through the silicon detectors and reconstructed calorimeter energy
- One project at University of Seoul is to classify quark versus gluon jets
  - Could be interesting in the context of new physics which couples to quarks, as jets in LHC more likely to be from gluon radiation
  - Gluons are more radiative, tend to be wider with more particles





Quark v Gluons: Model Comparison on Simulated Data



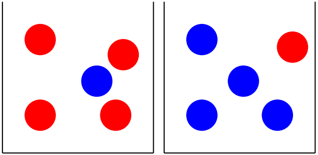


- Tested several CNN models with jet images versus a more traditional BDT made from several summary variables
- Showing AUC vs number of parameters for several models, and performance of ConvNet vs BDT versus jet energy
- Deep Learning gives improvement in q-g, several more complex models giving ~ same performance

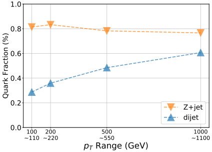
Lee et al., "Quark-Gluon Jet Discrimination Using Convolutional Neural Networks"
`https://doi.org/10.3938/jkps.74.219` arxiv : 2012.02531

Weakly-Supervised Learning Paradigm

Quark-bag Gluon-bag



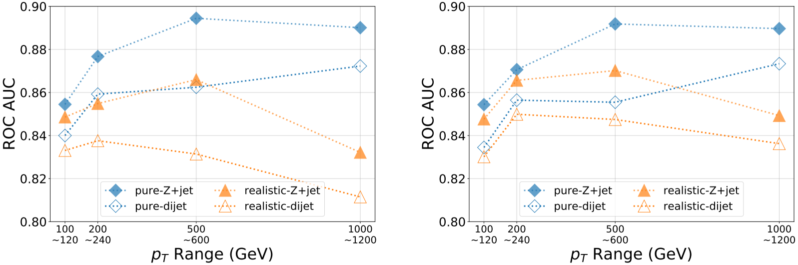
Predicted quark fraction



- QG jets difficult to simulate: try classification where instead of having quark/gluon labels, you have bag labels
  - One bag contains more gluons than quarks, the other more quarks
  - Training to classify the bags is the same as classifying q-g (assuming fraction of quarks is the only difference between the bags)
- Training labels based on bag distribution of quarks and gluons
  - We label 1 (0) when the jet is pulled from the gluon-enriched (quark-enriched) sample
- Suggests a data-driven approach to discrimination, as natural quark/gluon sources available
  - Z+jets jets are more likely to be quark-jets $\Longrightarrow$ quark bag
  - Dijet jets are more likely to be gluon-jets $\Longrightarrow$ gluon bag

"Weakly Supervised Classification in High Energy Physics" Dery et al. (1702.00414)
"Classification without labels", Metodiev et al. (arxiv:1708.02949)

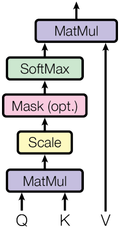
Weakly-supervised learning results

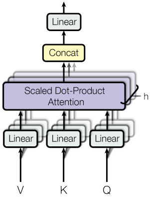


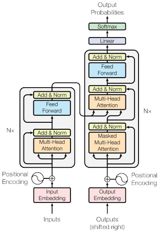
- Training done on a "realistic" simulated sample (WSL Z+jet versus dijet with predicted q-g fraction) and a "pure" sample (using true quark and gluon labels)
- At low p_T, performance of WSL near FSL, issue at high p_T (left)
    - Model seems to distinguish the difference in η distribution (# of jets in barrel vs endcap of detector), rather than pure quark/gluon differences
    - Can improve performance with an η cut during training (right)
- WSL works, but one must carefully check subtleties with this approach

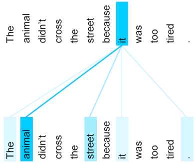
Lee et al, "Quark-Gluon Jet Discrimination Using Convolutional Neural Networks",
http://doi.org/10.3938/jkps.75.652 arxiv : 2012.02540

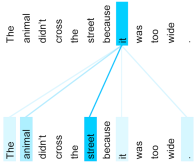
Self-Attention







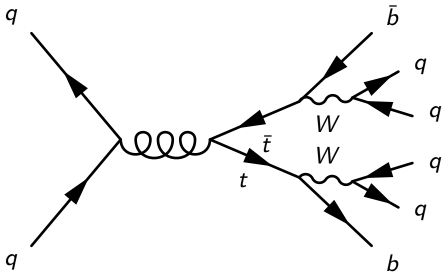


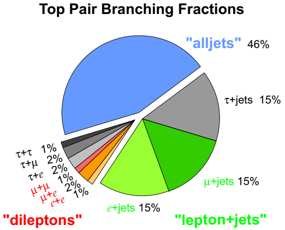


- The transformer model produces output sequence from input sequence
- Uses self-attention layers to model the dependencies between elements of the sequences: each word is trained to "attend" to related words
- The transformer model is at the heart of modern deep learning language processing
  - This paradigm replaces the previous recurrent models which processed single tokens at a time to produce a single context vector
  - Ever since it was argued that Attention is All You Need in 2017
- Achieved astonishing performance in applications like GPT-3

https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html

Self-Attention for Hadronic Top Pair Production





- Top quarks are interesting due to their very high mass (couplings to BSM?), and the only quark to decay before hadronization
- Top quarks are usually produced in pairs and decay in a cascade: $t \to Wq$, $W \to qq$ (or $\ell\nu$), two issues:
  - Jets come without labels (though can *b*-tag), so need to assign jets to reconstruct the top quark (and radiation can give additional jets)
  - Hard to distinguish from pure QCD background (= many jets, no ℓ)
- We created a Self-attention for Jet assignment (SaJa) network with jet inputs to assign jets to the top pair topology
  - Classify each jet as b from (anti-)top, q from (anti-)top, or other jets

Jet assignment loss function

- We take jet inputs to per-jet categories (avoid permutations):

$$f^{\theta} : \begin{pmatrix} \mathrm{x}^{(1)} \\ \vdots \\ \mathrm{x}^{(N)} \end{pmatrix} \rightarrow \begin{pmatrix} \hat{y}_{b_1}^{(1)} & \hat{y}_{W_1}^{(1)} & \hat{y}_{b_2}^{(1)} & \hat{y}_{W_2}^{(1)} & \hat{y}_{\mathrm{other}}^{(1)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{y}_{b_1}^{(N)} & \hat{y}_{W_1}^{(N)} & \hat{y}_{b_2}^{(N)} & \hat{y}_{W_2}^{(N)} & \hat{y}_{\mathrm{other}}^{(N)} \end{pmatrix}$$

- Hard to distinguish quarks from anti-quarks, so we create a loss function to account for this, which ignores top v. anti-top assignment

$$J(\theta) = \frac{1}{N} \sum_{j=1}^{N} \left(\min\left(\pi_{12}^{(j)}, \pi_{21}^{(j)}\right) + y_{\mathrm{other}}^{(j)} \log \hat{y}_{\mathrm{other}}^{(j)} \right)$$

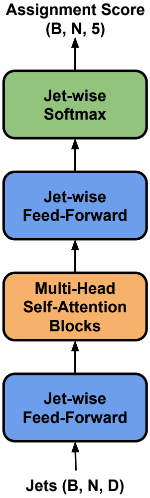
where

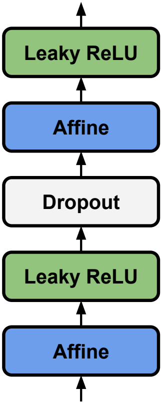
$$\pi_{\alpha\beta}^{(j)} = y_{b}^{(j)} \log \hat{y}_{b_\alpha}^{(j)} + y_{\bar{b}}^{(j)} \log \hat{y}_{b_\beta}^{(j)} + y_{W^+}^{(j)} \log \hat{y}_{W_\alpha}^{(j)} + y_{W^-}^{(j)} \log \hat{y}_{W_\beta}^{(j)}$$

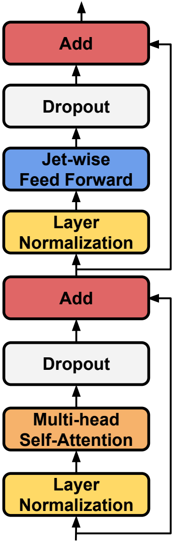
with α, $\beta \in \{1, 2\}$.

- Can extend this framework to other topologies (e.g. ttH, where $H \rightarrow b\bar{b}$) by using more categories

SaJa Network



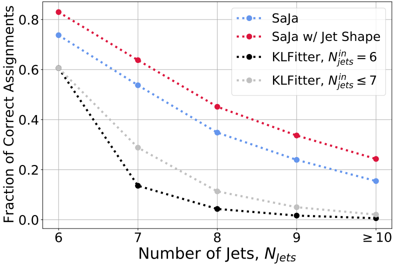


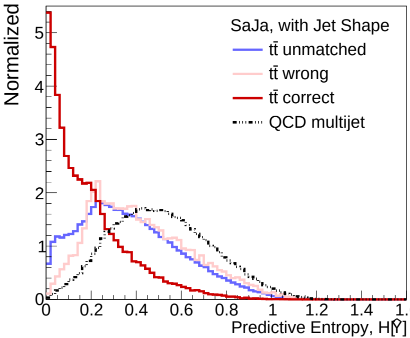


SaJa Jet-wise Attention block

- SaJa network consists of multi-head self-attention block wedged between jet-wise feed forward blocks
  - SA is the only place where the jets interact
- Trained on hadronic $t\bar{t}$ with fully resolved topology

Results from SaJa, QCD discrimination





- Greatly improve performance of standard likelihood-based assignment
- We also found the SaJa network *trained only to assign jets* can be used as a top-QCD classifier
  - Using jet-wise entropy $\mathbb{H}[\hat{Y}] = \frac{1}{N}\sum_{j=1}^{N}\left(-\sum_{c\in\text{classes}} \hat{y}_c^{(j)} \log \hat{y}_c^{(j)}\right)$
    - Entropy high for uncertain assignments, low when network is certain

Lee et al., "Zero-Permutation Jet-Parton Assignment using a Self-Attention Network"
https://arxiv.org/abs/2012.03542

Summary

- Presented a few analyses we've been working on at UoS
  - Comparison of contemporary deep learning models for jet id
  - Weakly supervised learning for data-driven deep learning
  - Self-attention for jet assignment
- Other topics we are looking at:
  - GAN for speeding up simulation (want at least as much simulation as data, major bottleneck)
  - Vision transformers, PointNet, CNN for particle identification in future colliders
  - Deep Learning on FPGAs for triggering physics events (as part of the 40 MHz $\rightarrow$ 1 kHz selection)
- Machine Learning taking an increasingly important role as dataset sizes and analysis difficulty increase